Online Versus Paper Test Comparability for the South Carolina English I and Physical Science End-of-Course Examinations Administered in Spring 2006

Pearson Educational Measurement

Report Prepared for the South Carolina Department of Education

September 14, 2006

Online Versus Paper Test Comparability for the South Carolina English I and Physical Science End-of-Course Examinations Administered in Spring 2006

Abstract

The purpose of this study was to evaluate comparability between student performances on end-of-course examinations taken in two different modes of administration. The first mode is the traditional paper and pencil version of the exam (hereon referred to as "paper" testing); while the second mode is the computer-based version (hereon referred to as "online" testing). Comparability of the two modes is evaluated for the spring 2006 administration of the South Carolina English I and Physical Science EOCEP tests. During this administration, the same test form was administered both in online and paper format for each course. This permitted comparisons to be made both at the overall score level and at the individual item level.

The study utilized an approach that considers score comparability in the context of test equating. Specifically, we drew samples of students testing online and matched them to the students taking the paper form using previous test scores, ethnicity, and gender. We then equated the online samples to the matched paper samples under the assumptions of a random groups design. We refer to this approach as matched samples comparability analyses (MSCA). The MSCA utilize a bootstrap method that is designed to estimate bootstrap standard errors of the equating to assist in interpreting differences between the online and paper score conversions (c.f., Kolen & Brennan, 2004, p. 232-235).

The samples analyzed included 4,160 students testing online and 30,703 students testing by paper for the English I EOCEP, and 3,893 students testing online and 26,131 students testing by paper for the Physical Science EOCEP during the spring 2006 administration. Each English I student testing online with a grade 8 PACT ELA score was matched to an English I student testing by paper with an identical profile of PACT ELA score level, ethnicity, and gender. Likewise, each Physical Science online student with a grade 8 PACT Science score was matched to a Physical Science paper student by PACT Science score level, ethnicity, and gender. To prevent empty cells, the grade 8

PACT scores for both ELA and Science were each ranked into 15 equal-sized groups and ethnicity was collapsed into black, white, and “other”. This resulted in 90 matching cells (15 ability groups × 3 ethnic groups × 2 gender groups) for each subject.

Results of the study indicated that the scores for the online and paper samples were comparable for both English I and Physical Science. In both subjects, analyses based on 100 replications indicated slightly higher performance for the online group, but the differences were one-half of a raw score point or less and were not statistically significant. The item-level analyses in both subjects flagged a small proportion of items with significant p-value differences. The differences, however, went in both directions and were consistent with what one might expect by chance.